Dual Optimal Resource Allocation for Heterogeneous Transmission in OFDMA Systems

Chunhui Liu[‡], Anke Schmeink[†] and Rudolf Mathar[‡]
[‡]Institute for Theoretical Information Technology,
[†]UMIC Research Centre,
RWTH Aachen University D-52056 Aachen, Germany
Email: liu@ti.rwth-aachen.de, schmeink@umic.rwth-aachen.de, mathar@ti.rwth-aachen.de

***Abstract*—This paper is concerned with dynamic resource allocation for heterogeneous downlink in orthogonal frequency division multiple access (OFDMA) systems. First, this paper formulates the resource allocation problem when one terminal may demand both non-real time (or best effort) transmission and real time transmission over the downlink simultaneously. We consider the constraint on the total transmission power of the base station for both types of services and the constraint on the rate requirement of each terminal for a real time transmission. To solve this problem, an efficient instantaneous per-symbol method with dual optimality is proposed based on the convex optimization framework. Compared to the exhaustive search, simulations show that the performance degradation of our method is limited to 0.002% of the optimal solution, when an OFDMA system consists of more than 16 subcarriers. Therefore, this paper illustrates that it is not computationally prohibitive to optimally solve the resource allocation problem for heterogeneous transmission over OFDMA downlink.**

I. Introduction

To efficiently combat the effects of frequency selective fading for high-speed wireless data transmission, the transmission band is divided into orthogonal subcarriers in OFDMA systems. If the bandwidth of subcarriers is sufficiently narrow, the subcarriers are subject to flat fading with different channel gains, detailed in [1]. To take advantage of the spatial diversity among terminals, OFDMA can allocate power and rate optimally to subcarriers depending on channel characteristics.

Services over OFDMA downlink can be generally classified into two groups according to different quality-of-service (QoS) requirements of the terminals. One is the rate-adaptive (RA) service, for which the throughput is maximized subject to the constraints on the fixed total transmission power at the base station and bit error rate (BER) required by each terminal. The other is the margin-adaptive (MA) service, for which the performance margin can be maximized while satisfying data rate and BER requirements of the terminals. The resource allocation problem for the RA services has been first studied in [2], [3]. Later, the minimum rate constraints or the proportional rate constraints are additionally considered in [4]–[6]. The methods in [7]–[10] respectively provide optimal, near-optimal and suboptimal solutions to the resource allocation problem for the MA services.

This work was supported by the UMIC Research Centre at RWTH Aachen University in Germany.

However, both groups of services may be required simultaneously over OFDMA downlink. A trade-off relationship exists between the subcarrier assignment and the power allocation for this case. If we assign more subcarriers to the MA services, more power could be kept for the RA services, but they would only have a few available subcarriers and may not have a large rate achievement, and vice versa. For the resource allocation for such heterogeneous transmission, the heuristic method from [11] gives a suboptimal solution by considering two consecutive steps. First, power and rate for the MA services are allocated, then, the remaining resources for the RA services. By using a similar procedure, [12] provides the suboptimal solution by performing greedy search over subcarriers. Dual optimal solutions may be obtained by the method suggested in [13]. Its complexity is sup-linearly increasing in the number of subcarriers, which is supposed to be very large in future communication systems.

In this paper, we investigate the resource allocation problem where the RA and MA services are demanded by terminals, even by one terminal, at the same time. We solve this problem by using the dual method, which has been applied to the resource allocation problem, e.g., [14]–[18]. We propose an efficient method to update the dual variables (Lagrange multipliers) to find the optimal values. Furthermore, we extend our method to solve the resource allocation problems for heterogeneous transmission, when the rates for the RA services are lower bounded or restricted by proportional rate constraints. The remainder of this paper is organized as follows. In Section II, the problem is formulated. In Section III, we use the duality theory to solve the resource allocation problem for heterogeneous transmission in OFDMA systems. Simulation results show that the duality gap becomes insignificantly when the number of subcarriers is large enough. Finally, the content of this paper is concluded.

II. Problem Formulation

We consider the downlink of an OFDMA system with one base station and K mobile terminals over N subcarriers. The transmission to different terminals is subject to independent frequency selective fading. Perfect channel knowledge is available at the base station and all terminals, so that they can perform resource allocation simultaneously and notification

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

of power and rate assignments is not necessary. A resource allocation scheme must satisfy two QoS requirements: rate and BER requirements. The uncoded modulation scheme of M-ary quadrature amplitude modulation (M-QAM) is employed.

In the considered system, a terminal k may require an MA service with an individual data rate R_k and $\mathrm{BER}_k^{(\mathrm{MA})}$, meanwhile, it may also require an RA service with $\mathrm{BER}_k^{(\mathrm{RA})}$. Generally, $\mathrm{BER}_k^{(\mathrm{RA})}$ for the RA service is lower than $\mathrm{BER}_k^{(\mathrm{MA})}$ for the MA service. Let $H_{k,n}^{(\mathrm{MA})}$ and $H_{k,n}^{(\mathrm{RA})}$ denote the channel gain-to-noise ratio (CNR) of subcarrier n multiplied with $\left[-1.5/\ln(5\mathrm{BER}_k^{(\mathrm{MA})})\right]$ and $\left[-1.5/\ln(5\mathrm{BER}_k^{(\mathrm{RA})})\right]$ for the MA and RA services of terminal k [19], respectively. The power and rate allocated to subcarrier n for terminal k are denoted by $P_{k,n}^{(\mathrm{RA})}$ and $r_{k,n}^{(\mathrm{RA})}$ for the RA service and $P_{k,n}^{(\mathrm{MA})}$ and $r_{k,n}^{(\mathrm{MA})}$ for the MA service. As shown in [1] the following relations hold:

$$
\begin{aligned}
r_{k,n}^{(\mathrm{MA})} &= \log_2(1 + P_{k,n}^{(\mathrm{MA})} H_{k,n}^{(\mathrm{MA})}),\\
r_{k,n}^{(\mathrm{RA})} &= \log_2(1 + P_{k,n}^{(\mathrm{RA})} H_{k,n}^{(\mathrm{RA})}).
\end{aligned}
$$

The aim is to maximize the total transmission rate for the RA services while satisfying the data rate and BER requirements for the MA services. The optimization problem can be stated as

$$
\begin{aligned}
&\text{maximize} && \sum_{k=1}^{K}\sum_{n=1}^{N} r_{k,n}^{(\mathrm{RA})} && (1)\\
&\text{subject to} && &&\\
&\text{C1:} && \sum_{n=1}^{N} r_{k,n}^{(\mathrm{MA})} \geq R_k, \qquad \forall\, k, &&\\
&\text{C2:} && \sum_{k=1}^{K}\sum_{n=1}^{N} (P_{k,n}^{(\mathrm{MA})} + P_{k,n}^{(\mathrm{RA})}) \leq P^{(\mathrm{tot})}, &&\\
&\text{C3:} && P_{k,n}^{(\mathrm{RA})} \geq 0, \qquad \forall\, k, \forall\, n, &&\\
& && P_{k,n}^{(\mathrm{MA})} \geq 0, \qquad \forall\, k, \forall\, n, &&\\
&\text{C4:} && \sum_{n=1}^{N} r_{k,n}^{(\mathrm{RA})} r_{l,n}^{(\mathrm{RA})} = 0, \ \ \forall\, k, l, k \neq l, &&\\
& && \sum_{n=1}^{N} r_{k,n}^{(\mathrm{MA})} r_{l,n}^{(\mathrm{MA})} = 0, \ \forall\, k, l, k \neq l, &&\\
& && \sum_{n=1}^{N} r_{k,n}^{(\mathrm{RA})} r_{l,n}^{(\mathrm{MA})} = 0, \qquad \forall\, k, l. &&
\end{aligned}
$$

The total transmission power of the base station is limited to $P^{(\mathrm{tot})}$. Constraint C4 illustrates that each subcarrier is not allowed to be shared among different terminals and services at the same time. To allow for theoretical analysis, continuous rates are considered throughout this paper. The sets of powers are denoted by $\{P_{k,n}^{(\mathrm{RA})}\} = \{P_{k,n}^{(\mathrm{RA})} | \forall\, k, n\}$ for the RA services and $\{P_{k,n}^{(\mathrm{MA})}\} = \{P_{k,n}^{(\mathrm{MA})} | \forall\, k, n\}$ for the MA services. Here, we does not consider the minimum or proportional rate constraints for the RA services, which may impair the generality of our solution. Later, they will be taken into account.

III. Dynamic Resource Allocation for Heterogeneous Transmission

To make the above problem feasible, the minimal sum of transmission power for the MA services must not be larger than $P^{(\mathrm{tot})}$, while the RA services are not required. Once the subcarrier assignment is determined for one service of one terminal, the power and rate allocation can be obtained by water-filling [1]. Multilevel water-filling from [7] can provide the optimal solution to the problem (1). However, $(2K)^N$ searches have to be performed to find the optimal combination of subcarrier assignments for different services and terminals. Its overall complexity is $\mathcal{O}(N(2K)^N)$, which makes it not implementable in practice. In this section, the dual method is applied to solve problem (1). It has been widely used on other problems of resource allocation for OFDMA systems or multiuser multiple-input and multiple-output systems, for example, [14]–[18]. However, to the best of our knowledge, it has not been applied to problem (1), where one terminal may require both types of services simultaneously.

A. Duality Gap for Nonconvex Optimization

Duality theory has been generally explained in [20] and specified for resource allocation problems in [16]. First, we only consider the first two constraints C1 and C2 but not C3 and C4. According to the Karush-Kuhn-Tucker conditions, the Lagrangian of problem (1) is defined as

$$
\begin{aligned}
&\mathcal{L}(\boldsymbol{\lambda}, \beta, \{P_{k,n}^{(\mathrm{RA})}\}, \{P_{k,n}^{(\mathrm{MA})}\})\\
&= -\sum_{k=1}^{K}\sum_{n=1}^{N} r_{k,n}^{(\mathrm{RA})} + \sum_{k=1}^{K} \lambda_k \Big(R_k - \sum_{n=1}^{N} r_{k,n}^{(\mathrm{MA})}\Big)\\
&\quad + \beta \left(\sum_{n=1}^{N}\sum_{k=1}^{K} (P_{k,n}^{(\mathrm{MA})} + P_{k,n}^{(\mathrm{RA})}) - P^{(\mathrm{tot})} \right), \qquad (2)
\end{aligned}
$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)^T$ and β are the Lagrange multipliers. Then, the dual optimization problem is

$$
\begin{aligned}
&\text{maximize} && \mathcal{D}(\boldsymbol{\lambda}, \beta) \qquad (3)\\
&\text{subject to} && \boldsymbol{\lambda} \succeq \mathbf{0}\\
& && \beta \geq 0,
\end{aligned}
$$

where the dual objective $\mathcal{D}(\boldsymbol{\lambda}, \beta)$ is defined as an unconstrained minimization of the Lagrangian, shown as

$$
\begin{aligned}
\mathcal{D}(\boldsymbol{\lambda}, \beta) &= \min_{\{P_{k,n}^{(\mathrm{RA})}\}, \{P_{k,n}^{(\mathrm{MA})}\}} \mathcal{L}(\boldsymbol{\lambda}, \beta, \{P_{k,n}^{(\mathrm{RA})}\}, \{P_{k,n}^{(\mathrm{MA})}\})\\
&= \sum_{k=1}^{K} \lambda_k R_k - \beta P^{(\mathrm{tot})}\\
&\quad + \min_{\{P_{k,n}^{(\mathrm{RA})}\}, \{P_{k,n}^{(\mathrm{MA})}\}} \sum_{n=1}^{N}\sum_{k=1}^{K} (\underbrace{\beta P_{k,n}^{(\mathrm{MA})} - \lambda_k r_{k,n}^{(\mathrm{MA})}}_{a_{k,n}}\\
&\quad + \underbrace{\beta P_{k,n}^{(\mathrm{RA})} - r_{k,n}^{(\mathrm{RA})}}_{b_{k,n}}).
\end{aligned}
$$

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

Since the power variables $\{P_{k,n}^{(\text{RA})}\}$ and $\{P_{k,n}^{(\text{MA})}\}$ are separable over subcarriers, we obtain that

$$\begin{aligned}\mathcal{D}(\boldsymbol{\lambda},\beta) &= \sum_{k=1}^{K}\lambda_k R_k - \beta P^{(\text{tot})} \\ &+ \sum_{n=1}^{N}\min_{\{P_{k,n}^{(\text{RA})}\},\{P_{k,n}^{(\text{MA})}\}}\sum_{k=1}^{K}(a_{k,n}+b_{k,n}).\end{aligned}$$

The objective of the primal problem is concave. Hence, the optimal duality gap of (1) is the sum of the primal and dual optimal solutions, and it must be non-negative. If we allow for time-sharing subcarriers among different services and terminals, the concavity of the objective function and the convexity of constraint C1 and C2 in (1) guarantee that the duality gap is equal to zero, which means strong duality, see *Theorem 1* in [16]. However, this is not permitted in the considered system while taking constraint C4 into account, so problem (1) is not a convex optimization problem and may not have strong duality, i.e., the duality gap may not be zero.

Then, due to C4, the power allocated to one subcarrier for at most one service of at most one terminal may be positive. Thus, the dual objective changes to

$$\begin{aligned}\mathcal{D}(\boldsymbol{\lambda},\beta) &= \sum_{k=1}^{K}\lambda_k R_k - \beta P^{(\text{tot})} \\ &+ \sum_{n=1}^{N}\min\left(\min_k a_{k,n}, \min_k b_{k,n}\right), \end{aligned} \tag{4}$$

where $\min(a,b)=a$ if $a\le b$, and vice versa.

Given $\boldsymbol{\lambda}$ and β, $a_{k,n}$ and $b_{k,n}$ are convex functions of $P_{k,n}^{(\text{MA})}$ and $P_{k,n}^{(\text{RA})}$, respectively. The minimum to $a_{k,n}$ and $b_{k,n}$ can be found by taking the derivative over $P_{k,n}^{(\text{MA})}$ and $P_{k,n}^{(\text{RA})}$, shown as

$$P_{k,n}^{(\text{MA})} = \left[\frac{\lambda_k}{\beta\ln 2} - \frac{1}{H_{k,n}^{(\text{MA})}}\right]^+, \tag{5}$$

$$P_{k,n}^{(\text{RA})} = \left[\frac{1}{\beta\ln 2} - \frac{1}{H_{k,n}^{(\text{RA})}}\right]^+, \tag{6}$$

where constraint C3 is considered and $[x]^+ = \max(x,0)$. The water levels are the terms $\lambda_k/(\beta\ln 2)$ and $1/(\beta\ln 2)$. Consequently, we have non-positive

$$a_{k,n} = \frac{\lambda_k}{\ln 2} - \frac{\beta}{H_{k,n}^{(\text{MA})}} - \lambda_k\log_2\left(\frac{\lambda_k H_{k,n}^{(\text{MA})}}{\beta\ln 2}\right), \tag{7}$$

$$b_{k,n} = \frac{1}{\ln 2} - \frac{\beta}{H_{k,n}^{(\text{RA})}} - \log_2\left(\frac{H_{k,n}^{(\text{RA})}}{\beta\ln 2}\right), \tag{8}$$

where $a_{k,n}$ is decreasing in λ_k, and both $a_{k,n}$ and $b_{k,n}$ are increasing in β.

By taking (7) and (8) to (4), subcarrier assignments to different services and terminals can be settled provided fixed $\boldsymbol{\lambda}$ and β. When $\boldsymbol{\lambda}$ and β changes, subcarrier assignments vary consequently. Obviously, this may cause a discrete change on the transmission power. Hence, the dual objective $\mathcal{D}(\boldsymbol{\lambda},\beta)$ may not be continuous at the optimal Lagrange multipliers $\boldsymbol{\lambda}^*$ and β^*. This does not satisfy the condition for strong duality, see *Theorem 2* in [16]. However, we will see that the duality gap becomes smaller for increasing N later on.

B. Efficient Updating Method

From the above subsection, the Lagrange multipliers $\boldsymbol{\lambda}$ and β are the key to find the dual optimal solution to (1). Compared to the subgradient method, as an example for the cutting-plane method, the ellipsoid method can update $\boldsymbol{\lambda}$ and β more efficiently. First, it localizes the optimal values $\boldsymbol{\lambda}^*$ and β^* in a closed and bounded region. Then, at the center of the region, evaluation of the gradient or subgradient of $\mathcal{D}(\boldsymbol{\lambda},\beta)$ and shrinking the region are iteratively performed. The region continues shrinking until it converges to $\boldsymbol{\lambda}^*$ and β^*.

Given $\boldsymbol{\lambda}$ and β, $\{P_{k,n}^{(\text{MA})^*}\}$ and $\{P_{k,n}^{(\text{RA})^*}\}$ minimize (2) to $\mathcal{D}(\boldsymbol{\lambda},\beta) = \mathcal{L}(\boldsymbol{\lambda},\beta,\{P_{k,n}^{(\text{RA})^*}\},\{P_{k,n}^{(\text{MA})^*}\})$. For any $\boldsymbol{\lambda}' \succeq \mathbf{0}$ and $\beta' \ge 0$, it follows that

$$D(\boldsymbol{\lambda}',\beta) \le D(\boldsymbol{\lambda},\beta) + \sum_{k=1}^{K}(\lambda'_k - \lambda_k)\left(R_k - \sum_{n=1}^{N} r_{k,n}^{(\text{MA})^*}\right),$$

$$D(\boldsymbol{\lambda},\beta') \le D(\boldsymbol{\lambda},\beta) + (\beta'-\beta)\left(\sum_{n=1}^{N}\sum_{k=1}^{K}(P_{k,n}^{(\text{MA})^*} + P_{k,n}^{(\text{RA})^*}) - P^{(\text{tot})}\right),$$

see *Proof* in [17].

If we perform the ellipsoid method over $\boldsymbol{\lambda}$ and β, the region containing their optimal values has to be found as earlier mentioned. Due to the heterogeneous constraints C1 and C2, the initialization for such a region given in [16] does not fit our problem and it has the complexity of $\mathcal{O}(KN)$. To simply and intuitively initialize the ellipsoid method, we perform it over the water levels and make the following substitution:

$$y = \frac{1}{\beta\ln 2} \quad \text{and} \quad x_k = \frac{\lambda_k}{\beta\ln 2},\ \forall\, k.$$

From (5) and (6), on one side, the water levels $\boldsymbol{x}$ and y must not be smaller than the reciprocals of the corresponding largest CNRs. On the other side, for the MA service of terminal k, the highest water level could be achieved when the required rate R_k is allocated only to the subcarrier with the smallest CNR. The water level y for RA services could reach the maximal value, if all transmission power $P^{(\text{tot})}$ is allocated only to the subcarrier with the smallest CNR among all terminals. Hence, the following inequalities always hold

$$\min_n \frac{1}{H_{k,n}^{(\text{MA})}} \le \quad x_k \quad \le 2^{R_k}\max_n \frac{1}{H_{k,n}^{(\text{MA})}},\ \forall\, k \tag{9}$$

$$\min_{n,k} \frac{1}{H_{k,n}^{(\text{RA})}} \le \quad y \quad \le P^{(\text{tot})} + \max_{n,k}\frac{1}{H_{k,n}^{(\text{RA})}}, \tag{10}$$

which defines the initial region containing the optimal water levels for the ellipsoid method.

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

After such a replacement, the Lagrangian (2) becomes

$$\begin{aligned}&\mathcal{L}_s(\boldsymbol{x}, y, \{P_{k,n}^{(\text{RA})}\}, \{P_{k,n}^{(\text{MA})}\})\\ &= -\sum_{k=1}^{K}\sum_{n=1}^{N} r_{k,n}^{(\text{RA})} + \sum_{k=1}^{K}\frac{x_k}{y}\underbrace{(R_k - \sum_{n=1}^{N} r_{k,n}^{(\text{MA})})}_{c_k}\\ &\quad + \frac{1}{y\ln 2}\underbrace{\left(\sum_{n=1}^{N}\sum_{k=1}^{K}(P_{k,n}^{(\text{MA})} + P_{k,n}^{(\text{RA})}) - P^{(\text{tot})}\right)}_{c_{K+1}},\end{aligned}$$

where $\boldsymbol{x} = (x_1, \ldots, x_k)^T$.

Given $\boldsymbol{x}$ and y, $\{c_k^*\} = \{c_k^* | \forall\, k \in \{1, \ldots, K+1\}\}$ minimize (2) to $\mathcal{D}_s(\boldsymbol{x}, y) = \mathcal{L}_s(\boldsymbol{x}, y, \{c_k^*\})$. On one hand, for any $\boldsymbol{x}' \succeq \mathbf{0}$, we have

$$\begin{aligned}\mathcal{D}_s(\boldsymbol{x}', y) &\leq \mathcal{L}_s(\boldsymbol{x}', y, \{c_k^*\})\\ &= \mathcal{L}_s(\boldsymbol{x}, y, \{c_k^*\}) + \sum_{k=1}^{K}\Delta x_k \frac{c_k^*}{y},\end{aligned}$$

where $\Delta x_k = x_k' - x_k$. On the other hand, for any $y' \geq 0$, it follows that

$$\begin{aligned}\mathcal{D}_s(\boldsymbol{x}, y') &\leq \mathcal{L}_s(\boldsymbol{x}, y', \{c_k^*\})\\ &= \mathcal{L}_s(\boldsymbol{x}, y, \{c_k^*\}) + \Delta y\, d,\end{aligned}$$

where $\Delta y = y' - y$ and

$$d = -\frac{1}{y^2 + y\Delta y}\left(\sum_{k=1}^{K} x_k c_k^* + \frac{1}{\ln 2}c_{K+1}^*\right).$$

If Δy is small, it holds that

$$d = -\frac{1}{y^2}\left(\sum_{k=1}^{K} x_k c_k^* + \frac{1}{\ln 2}c_{K+1}^*\right).$$

Since $\mathcal{L}_s(\boldsymbol{x}, y, \{P_{k,n}^{(\text{RA})}\}, \{P_{k,n}^{(\text{MA})}\})$ is convex, its subgradient is

$$\boldsymbol{g} = -(\frac{c_1^*}{y}, \ldots, \frac{c_K^*}{y}, d)^T.$$

With (9), (10) and the above subgradient, the ellipsoid method can be performed to solve our resource allocation problem.

C. Complexity Analysis and Extension

The initial region, defined by (9) and (10), is not so tight as the ones in [15]–[18]. However, this only impairs the converging speed of our method slightly due to the exponential shrinking behavior of the bounded region in the ellipsoid method as earlier illustrated. The number of iterations while performing the ellipsoid method is proportional to i^2, where i denotes the number of variables [21]. In each iteration, (7) and (8) must be calculated $2KN$ times. The overall complexity of the proposed method is $\mathcal{O}(2KN(K+1)^2) \ll \mathcal{O}(2K)^N$. This means that our method can be used in the small wireless networks, where a small number of terminals exist or the channel does not vary very fast. Further, our method provides a reference for designing low-complexity heuristic methods solving our resource allocation problem, where the exhaustive search cannot be used for large systems.

In previous works [11]–[13] on the heterogeneous transmission, it is assumed that one terminal requires only one service at a specific period of time. To fit such a problem, we set $R_k = 0$ for $k \in \mathcal{R}$, where $\mathcal{R}$ is the set of terminals only requiring the RA services, and set $P_{k,n}^{(\text{RA})} = 0$ for $k \in \mathcal{M}$, where $\mathcal{M}$ is the set of terminals only requiring the MA services. Then, the dual optimal solution can be acquired by performing the ellipsoid method over the $|\mathcal{M}| + 1$ water levels.

When the minimum rate constraints $\sum_{n=1}^{N} r_{k,n}^{(\text{RA})} \geq M_k, \forall\, k$ for RA services are additionally considered in problem (1), the term $\sum_{k=1}^{K}\gamma_k(M_k - \sum_{n=1}^{N} r_{k,n}^{(\text{RA})})$ would be added to the Lagrangian (2). The rate for the RA service of terminal k is lower bounded by M_k. The new Lagrange multipliers are $\{\gamma_1, \ldots, \gamma_K\}$. The duality theory, the substitution and the initialization of the search region described in the above two subsections can still be used to perform the ellipsoid method in order to find the dual optimal solution. Then, the complexity increases to $\mathcal{O}(2KN(2K+1)^2)$.

If the proportion of the achieved rates for the different RA services must be equal to $\phi_1 : \ldots : \phi_K$, the objective of problem (1) can be rewritten as $W\sum_{k=1}^{K}\phi_k$ and constraints $\sum_{n=1}^{N} r_{k,n}^{(\text{RA})} = W\phi_k$ are considered. Non-negative-definite W is the normalizer of the achieved rates for the RA services and is monotonically increasing in $P^{(\text{tot})}$. It can be upper bounded by the following procedure intuitively. First, all subcarriers are assigned to each RA service. Then, W must be smaller than $\max_k\left(\sum_{n=1}^{N} r_{k,n}^{(\text{RA})}/\phi_k\right)$. The complexity of this procedure is $\mathcal{O}(KN)$. The bisection method can be used to find the dual optimal W. Given W, the duality theory and the ellipsoid method can be used to solve the following problem [17]

$$\begin{aligned}\text{minimize} \quad & \sum_{k=1}^{K}\sum_{n=1}^{N}(P_{k,n}^{(\text{MA})} + P_{k,n}^{(\text{RA})})\\ \text{subject to} \quad & \\ & \sum_{n=1}^{N} r_{k,n}^{(\text{MA})} \geq R_k, \qquad \forall\, k,\\ & \sum_{n=1}^{N} r_{k,n}^{(\text{RA})} = W\phi_k, \qquad \forall\, k,\\ & P_{k,n}^{(\text{RA})} \geq 0, \qquad \forall\, k, \forall\, n,\\ & P_{k,n}^{(\text{MA})} \geq 0, \qquad \forall\, k, \forall\, n,\\ & \sum_{n=1}^{N} r_{k,n}^{(\text{RA})} r_{l,n}^{(\text{RA})} = 0, \quad \forall\, k, l, k \neq l,\\ & \sum_{n=1}^{N} r_{k,n}^{(\text{MA})} r_{l,n}^{(\text{MA})} = 0, \quad \forall\, k, l, k \neq l,\\ & \sum_{n=1}^{N} r_{k,n}^{(\text{RA})} r_{l,n}^{(\text{MA})} = 0, \qquad \forall\, k, l.\end{aligned}$$

When $|P^{(\text{tot})} - \sum_{n=1}^{N}\sum_{k=1}^{K}(P_{k,n}^{(\text{MA})} + P_{k,n}^{(\text{RA})})| \leq \epsilon, \epsilon \in (0,1)$ is satisfied, the bisection search finishes.

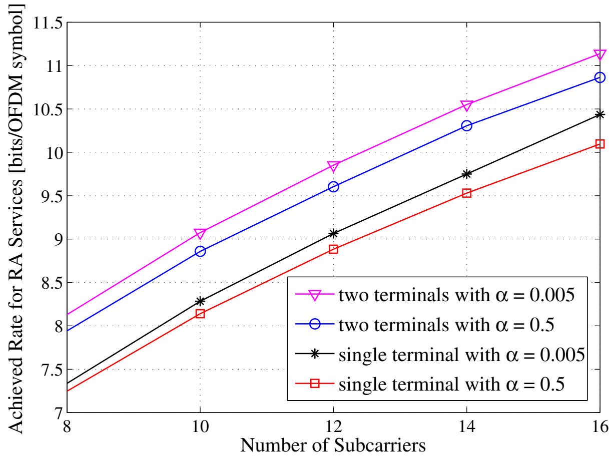

Fig. 1. Rate achievement vs. number of subcarriers.

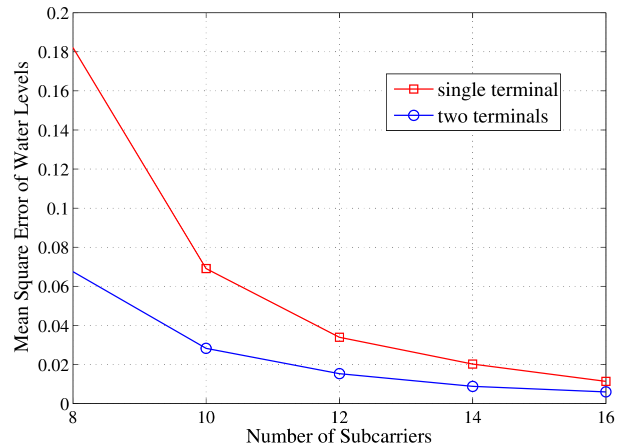

Fig. 2. MSE of water levels vs. number of subcarriers with $\alpha = 0.5$.

IV. SIMULATION RESULTS

In this section, our method is compared to the multilevel water-filling from [7]. The expected rate achievement, the duality gap and the mean square error (MSE) of the water levels are recorded. The frequency selective channel is modeled as consisting of $Q{=}8$ taps with an exponentially decaying profile, given by the impulse response $h(t){=}\sum_{q=0}^{Q-1} e^{-q\alpha} h_q \delta(t-qT)$. The taps h_q are jointly independent, circular symmetric, complex Gaussian distributed with zero mean and variance 1. Time T is the tap interval, q the tap index, α the positive decaying exponent, and t the time index. The expected CNR on each subcarrier is set to 5 dB. Due to the high complexity of the multilevel water-filling, we only consider a small OFDMA system with 8 to 16 subcarriers to keep $N \geq Q$. MA services with $\mathrm{BER}^{(\mathrm{MA})}{=}2.55{\times}10^{-3}$ in Table I and RA services with $\mathrm{BER}^{(\mathrm{RA})} = 2.63 \times 10^{-4}$ are required. The total transmission power at the base station is limited to 12 dB. A number of 25 000 channel samples are generated for each simulation.

Fig. 1 gives the rate achievement for the RA service with different α in two cases. One is that both services are required by a single terminal, and the other is that these two services are demanded by two terminals separately. Stronger frequency selectivity, due to a smaller α, enlarges the diversity of CNRs across subcarriers and among terminals. Consequently, the rate achievement is enhanced. Given the same α, larger channel diversity among the two terminals leads to a larger rate achievement compared to the single-terminal case.

Although the MSE of water levels for the case of two terminals is smaller than that for the single-terminal case, shown in Fig. 2, the duality gaps normalized by the achieved rates have an inverse behavior drawn in Fig. 3. To explain this, we roughly divide subcarriers into a good set with higher CNRs and a bad set with lower CNRs for each service. For the case of a single terminal, two services have exactly the same good and bad sets, which means that there is no channel diversity between services. The gaps between the optimal subcarrier assignment and other subcarrier assignments must be very large. In other words, relatively good suboptimal subcarrier assignments are very few. Even given bad suboptimal water levels, the best subcarrier assignment may still be chosen.

On the contrary, when the RA and MA services are separately demanded by two terminals, services have different good sets and bad sets of subcarriers due to the channel diversity between terminals. This results in more relatively good suboptimal subcarrier assignments. Even when the water levels provided by the ellipsoid method are only slightly different from the optimal ones, the optimal subcarrier assignments may not be derived. From Fig. 3, it can be deduced that more than 99.998% of the optimal rate would be allocated while $N > 16$ holds in our simulation.

TABLE I
MA SERVICES IN THE SIMULATION SYSTEM

type	Proportion	Rate [bits/OFDM symbol]
Video service	10%	4
Data service	40%	2
Audio service	50%	1

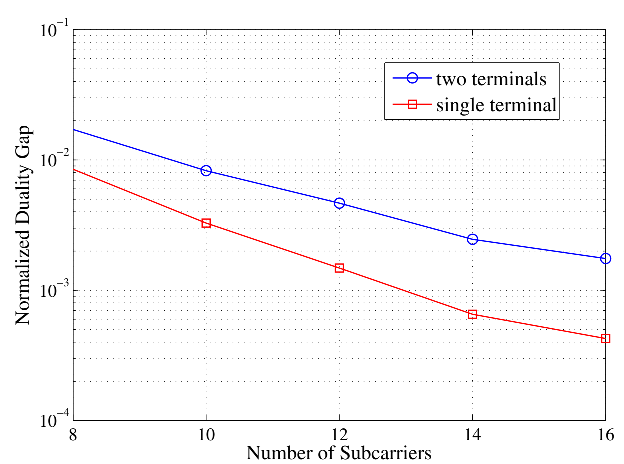

Fig. 3. Duality gap vs. number of subcarriers with $\alpha = 0.5$.

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

V. Conclusion

In this work, the resource allocation problem for heterogeneous transmission in OFDMA systems has been generally formulated. To solve this problem, a dual optimal method has been proposed by using the duality theory and the ellipsoid method. Its complexity is linearly increasing in the number of subcarriers and is much lower than that of the exhaustive search. Moreover, it has been extended to provide the dual optimal solutions, when the minimum rate constraints or the proportional rate constraints for the RA services are additionally considered. It offers a bench mark for evaluating low-complexity heuristic methods giving suboptimal solutions to our problem, when the high-complexity exhaustive search cannot be used. It is possible to use the proposed method in the OFDMA systems with a small number of terminals or with slow fading channels. In our simulation, 99.998% of the optimal rate could be achieved when systems consist of more than 16 subcarriers. In practice, a large number of subcarriers would make the solution given by our method even closer to the optimal one.

References

[1] Y. Li and G. L. Stueber, Eds., *Orthogonal Frequency Division Multiplexing for Wireless Communications*. Springer, 2005.

[2] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE Trans. Commun.*, vol. 21, pp. 171–178, 2003.

[3] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proc. IEEE VTC*, vol. 2, May 2000, pp. 1085–1089.

[4] M. Mehrjoo, M. Somayeh, and X. Shen, "A new modeling approach for utility-based resource allocation in OFDM networks," in *Proc. IEEE ICC*, May 2008.

[5] W. Xu, C. Zhao, P. Zhou, and Y. Yang, "Efficient adaptive resource allocation for multiuser OFDM systems with minimum rate constraints," in *Proc. IEEE ICC*, june 2007.

[6] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2726–2737, Nov. 2005.

[7] L. M. C. Hoo, B. Halder, J. Tellado, and J. M. Cioffi, "Multiuser transmit optimization for multicarrier broadcast channels: asymptotic FDMA capacity region and algorithms," *IEEE Trans. Commun.*, vol. 52, pp. 922–930, June 2004.

[8] D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 1150– 1158, Nov. 2003.

[9] C. Y. Wong, R. Cheng, K. Letaief, and R. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 1747–1758, Oct. 1999.

[10] C. Liu, A. Schmeink, and R. Mathar, "Power allocation for broadcasting in multiuser OFDM systems with sublinear complexity," to appear in *Proc. IEEE ICC*, June 2009.

[11] W. Wang, K. C. Hwang, K. B. Lee, and S. Bahk, "Resource allocation for heterogeneous services in multiuser OFDM systems," in *Proc. IEEE GLOBECOM*, Dec. 2004.

[12] V. Majjigi, R. Agarwal, R. Vannithamby, and J. Cioffi, "Persistent resource allocation in OFDMA networks for real-time and non-real-time traffic," in *Proc. IEEE ACSSC*, Nov. 2007.

[13] S. Bashar and Z. Ding, "Efficient algorithms for resource allocation in heterogeneous OFDMA networks," in *Proc. IEEE GLOBECOM*, Dec. 2008.

[14] I. C. Wong and B. L. Evans, "Optimal downlink OFDMA resource allocation with linear complexity to maximize ergodic rates," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 962–971, March 2008.

[15] R. Zhang, "Optimal dynamic resource allocation for multi-antenna broadcasting with heterogeneous delay-constrained traffic," *IEEE J. Sel. Topics. Signal Process*, vol. 2, pp. 243–255, April 2008.

[16] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, pp. 1310–1322, 2006.

[17] K. Seong, M. Mohseni, and J. M. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *Proc. IEEE ISIT*, July 2006.

[18] M. Mohseni, R. Zhang, and J. M. Cioffi, "Optimized transmission for fading multiple-access and broadcast channels with multiple antennas," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 1627–1639, Aug. 2006.

[19] A. J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 1218–1230, Oct 1997.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[21] H. T. Jongen, K. Meer, and E. Triesch, *Optimazation Theory*. Kluwer Academic Publishers, 2004.

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.